制御マルコフ連鎖上での閾値確率最適化の方法 (不確実性の下での数理モデルの構築と最適化)

AUTHOR(S):

植野, 貴之; 岩本, 誠一

CITATION:

植野, 貴之 ...[et al]. 制御マルコフ連鎖上での閾値確率最適化の方法 (不確実性の下での数理モデルの構築と最適化). 数理解析研究所講究録 2001, 1194: 24-32

ISSUE DATE:

2001-03

URL:

http://hdl.handle.net/2433/64823

数理解析研究所講究録
1194巻 2001年 24-32

制御マルコフ連鎖上での閾値確率最適化の方法

九大大学院経済学研究科　植野貴之 (Takayuki Ueno)
Graduate School of Economics, Kyushu University
九大大学院経済学研究院　岩本誠一 (Seiichi Iwamoto)
Faculty of Economics, Kyushu University

1　はじめに

本論文では、有限段制御マルコフ連鎖において最小型評価値が所定の値以上になる閾値確率を最大化する問題を考える。この閾値確率制御問題に対して５つの方法 ―― （１）パラメトリック法、（２）全履歴法、（３）マルコフ法、（４）多段確率決定樹表、（５）政策列挙法 ―― によって共通の最適解が得られることを示す。このうち、（１）、（２）、（３）が動的計画法である。（４）は（２）全履歴法を図と表によってビジュアル化したものである。（５）ではマルコフ政策による閾値確率ベクトルをすべて列挙して、１枚の表にまとめている。この表は、各マルコフ政策による閾値確率ベクトルを求めて構成されるが、（４）多段確率決定樹表から選び出したものでもある。

いわゆる加法型評価の期待値最適化問題ではマルコフ政策クラスの中で最適政策が得られる。すなわち、マルコフ政策は十分である ([3],[4])。これに対して、「閾値確率」制御問題ではより広い一般政策クラスに最適解が得られる。すなわち、加法型評価系に対する閾値確率制御問題ではマルコフ政策は十分でない ([5],[6],[8])。（１）パラメトリック法では「拡大」マルコフ政策クラスで最適化し、（２）全履歴法 および （４）多段確率決定樹表 では原始政策クラスで最適化する。いずれにおいても得られた最適政策を一般政策クラスに圧縮して、求める最適政策が得られる。さらに、（３）マルコフ法では、「最小型評価に対する閾値確率の特性」を用いて、マルコフ政策クラスの中で最適化しても最適政策が得られることを示す。

さらに、３状態２決定２段モデルに対して５つの方法によって具体的に最適解を求め、一致することが示される。

本論文では、有限段マルコフ連鎖において最小型評価系に対する閾値確率制御問題を３つの政策の視点から最適政策を中心に究明している。無限段マルコフ決定過程においては、加法型評価系の閾値確率について、本論文と異なった方法で研究されている ([2],[10])。

2　閾値確率制御問題

本節では、不確実性の下で最小型評価の閾値確率制御を考える。以後全体を通して次のデータが与えられているものとする：

(1)　$N \geq 2$ は**段の総数** (total number of stage) を表す正整数
(2)　$X = \{s_1, s_2, \ldots, s_p\}$ は有限**状態空間** (state space)
(3)　$U = \{a_1, a_2, \ldots, a_k\}$ は有限**決定空間** (action space)
(4)　$r_n : X \times U \to R^1$ は**第 n 利得関数** (n-th reward function)　$(0 \leq n \leq N-1)$
　　$r_N : X \to R^1$ は**終端利得関数** (terminal reward function)
(5)　$p = \{p(y|x,u)\}$ は**マルコフ推移法則** (Markov transition law)

: $p(y|x,u) \geq 0 \quad \forall (x,u,y) \in X \times U \times X$

$\sum_{y \in X} p(y|x,u) = 1 \quad \forall (x,u) \in X \times U$

(6) $\pi = \{\pi_0, \pi_1, \ldots, \pi_{N-1}\}$ は**マルコフ政策** (Markov policy)

: $\pi_0 : X \to U, \quad \pi_1 : X \to U, \quad \ldots, \quad \pi_{N-1} : X \to U$

マルコフ政策の全体を Π で表わす。

(6)′ $\sigma = \{\sigma_0, \sigma_1, \ldots, \sigma_{N-1}\}$ は**一般政策** (general policy)

: $\sigma_0 : X \to U, \quad \sigma_1 : X \times X \to U, \quad \ldots, \quad \sigma_{N-1} : X \times \cdots \times X \to U$

一般政策の全体を Π_g で表わす。

(6)″ $\mu = \{\mu_0, \mu_1, \ldots, \mu_{N-1}\}$ は**原始政策** (primitive policy)

: $\mu_0 : X \to U, \quad \mu_1 : X \times U \times X \to U, \quad \ldots, \quad \mu_{N-1} : X \times U \times X \times U \times \cdots \times U \times X \to U$

原始政策の全体を Π_p で表わす。

このとき、次の問題を考える：

$$\begin{array}{ll} \text{Max} & P^{\sigma}_{x_0}(r_0 \wedge r_1 \wedge \cdots \wedge r_{N-1} \wedge r_N \geq c) \\ \text{s.t.} & (\text{i})_n \quad X_{n+1} \sim p(\cdot|x_n, u_n) \\ & (\text{ii})_n \quad u_n \in U \quad n = 0, 1, \ldots, N-1 \end{array} \tag{1}$$

ただし $r_n = r_n(X_n, U_n)$, $r_N = r_N(X_N)$ で、$Y \sim p(\cdot|x,u)$ は現時刻の状態が x、決定が u であるとき、次の時刻で状態 y へ確率 $p(y|x,u)$ で推移することをあらわす。また $P^{\sigma}_{x_0}$ は条件付き確率 $p(\cdot|\cdot,\cdot)$、政策 $\sigma = \{\sigma_0, \sigma_1, ..., \sigma_{N-1}\}$ 及び初期状態 $x_0 \in X$ に依存して定まる全履歴空間 $X \times U \times X \times U \times \cdots \times U \times X$ 上の確率測度を表す。したがって、意志決定者が一般（現在までの状態列に依存する）政策 σ を採用すると、最大化問題 (1) の閾値確率は「部分」多重和

$$\begin{aligned} & P^{\sigma}_{x_0}(r_0 \wedge r_1 \wedge \cdots \wedge r_{N-1} \wedge r_N \geq c) \\ = & \sum_{(x_1, x_2, \ldots, x_N) \in (*)} \sum \cdots \sum p(x_1|x_0, u_0) p(x_2|x_1, u_1) \cdots p(x_N|x_{N-1}, u_{N-1}) \end{aligned} \tag{2}$$

で表わされる。ただし、多重和をとる領域 $(*)$ は

$$r_0(x_0, u_0) \wedge r_1(x_1, u_1) \wedge \cdots \wedge r_{N-1}(x_{N-1}, u_{N-1}) \wedge r_N(x_N) \ \geq \ c \tag{3}$$

を満たす $(x_1, x_2, \ldots, x_N) \in X \times X \times \cdots \times X$ 全体である。ここに、式 (2),(3) における決定列 $\{u_0, u_1, \cdots, u_{N-1}\}$ は一般政策 $\sigma = \{\sigma_0, \ldots, \sigma_{N-1}\}$ を通して定まっている：

$$u_0 = \sigma_0(x_0), \ u_1 = \sigma_1(x_0, x_1), \ \ldots, \ u_{N-1} = \sigma_{N-1}(x_0, x_1, \ldots, x_{N-1}).$$

一般に、確率変数 Y が c 以上になる確率 $P(Y \geq c)$ は、数直線 R^1 上の区間 $[c, \infty)$ の定義関数

$$\psi(y) := 1_{[c,\infty)}(y) := \begin{cases} 1 & y \geq c \\ 0 & \text{その他} \end{cases}$$

を通した確率変数 $\psi(Y)$ の期待値 $E[\psi(Y)]$ で表わされる：

$$P(Y \geq c) = E[\psi(Y)].$$

このことに注意すると、一般問題 (1) の閾値確率は定義関数 $\psi = \psi(y)$ を通した期待値になる：

$$P^{\sigma}_{x_0}(r_0 \wedge \cdots \wedge r_{N-1} \wedge r_N \geq c) \ = \ E^{\sigma}_{x_0}[\psi(r_0 \wedge \cdots \wedge r_{N-1} \wedge r_N)].$$

すなわち、「部分」多重和は定義関数を通した「全」多重和に等しい。

3 拡大マルコフ政策クラス

問題 (1) に対して、**過去値集合列** $\{\Lambda_n\}$ を

$$\begin{aligned}\Lambda_0 &:= \{\lambda_0\} \quad \lambda_0 \text{ は、} r_n, r_N \text{の取り得る最大値}\\ \Lambda_n &:= \{\lambda_n \,|\, \lambda_n = r_0(x_0,u_0)\wedge\cdots\wedge r_{n-1}(x_{n-1},u_{n-1}),\\ &\qquad (x_0,u_0,\ldots,x_{n-1},u_{n-1}) \in X\times U\times\cdots\times X\times U\}\\ &\qquad\qquad n=1,\ldots,N\end{aligned}$$

で定義すると、相隣る過去値集合 $\{\Lambda_{n-1},\ \Lambda_n\}$ 間に次の前向きの再帰式が成り立つ：

補題 3.1 (前向きの再帰式)

$$\begin{aligned}&\Lambda_0 = \{\lambda_0\}\\ &\Lambda_n = \{\lambda\wedge r_{n-1}(x,u)\,|\,\lambda\in\Lambda_{n-1},\ (x,u)\in X\times U\} \qquad n=1,2,\ldots,N.\end{aligned}$$

さらに、第 n 段までの**過去値確率変数** $\tilde{\Lambda}_n$：

$$\begin{aligned}\tilde{\Lambda}_0 &:= \lambda_0 \quad \text{ただし } \lambda_0 \text{ は十分大きな定数}\\ \tilde{\Lambda}_n &:= r_0(X_0,U_0)\wedge\cdots\wedge r_{n-1}(X_{n-1},U_{n-1})\end{aligned}$$

を導入すると、拡大状態空間列 $\{X\times\Lambda_n\}_0^N$ 上の**終端型期待値**評価問題

$$\begin{aligned}\text{Max}\quad & \tilde{E}^{\gamma}_{y_0}[\psi(\tilde{\Lambda}_N\wedge r_N(X_N))]\\ \text{s.t.}\quad & (\text{i})_n,\ (\text{ii})_n \quad n=0,1,\ldots,N-1\\ & (\text{i})'_n \quad \tilde{\Lambda}_{n+1} = \tilde{\Lambda}_n\wedge r_n(X_n,U_n)\end{aligned}$$

が考えられる。ただし、 $y_0=(x_0;\lambda_0)$. ここに $\tilde{E}^{\gamma}_{y_0}$ は、初期状態 y_0 、拡大マルコフ政策 γ および新マルコフ推移法則 q によって拡大状態空間列上に定まる確率測度 $\tilde{P}^{\gamma}_{y_0}$ に基づく期待値作用素である。

この終端型問題に対して、拡大状態 $y_n=(x_n;\lambda_n)$ から始まる部分問題

$$\begin{aligned}\text{Max}\quad & \tilde{E}^{\gamma}_{y_n}[\psi(\tilde{\Lambda}_N\wedge r_N(X_N))]\\ \text{s.t.}\quad & (\text{i})_m,\ (\text{ii})_m,\ (\text{i})'_m \quad m=n,\ldots,N-1\end{aligned}$$

の最大値を $u^n(y_n)$ とすると、次の再帰式が成り立つ ([6],[7])：

定理 3.1 (後向きの再帰式)

$$\begin{aligned}u^N(x;\lambda) &= \psi(\lambda\wedge r_N(x)) \qquad x\in X,\ \lambda\in\Lambda_N\\ u^n(x;\lambda) &= \max_{u\in U}\sum_{y\in X} u^{n+1}(y;\lambda\wedge r_n(x,u))p(y|x,u)\\ &\qquad\qquad x\in X,\ \lambda\in\Lambda_n,\ 0\le n\le N-1.\end{aligned}$$

4 原始政策クラス

さらに、(1) に対しては、原始（全履歴に依存する）政策クラス上の問題が考えられる：

$$\begin{aligned}\text{Max}\quad & P^{\mu}_{x_0}(r_0\wedge\cdots\wedge r_{N-1}\wedge r_N\ge c)\\ \text{s.t.}\quad & (\text{i})_n,\ (\text{ii})_n \quad n=0,1,\ldots,N-1\end{aligned}$$

これに対しては、後向きの再帰式が成り立つ ([9])：

定理 4.1 （後向きの再帰式）

$$w_n(h) = \underset{u\in U}{\mathrm{Max}} \sum_{y\in X} w_{n+1}(h,u,y)p(y|x,u) \qquad h \in H_n, \quad 1 \le n \le N-1$$
$$w_{N+1}(h) = \psi(r_0(x_0,u_0) \wedge \cdots \wedge r_{N-1}(x_{N-1},u_{N-1}) \wedge r_N(x_N)) \qquad h \in H_N.$$

5 マルコフ政策クラス

この節では、最小型評価に対する閾値確率の特性を用いて、マルコフ政策クラスの中で最適化しても最適政策が得られることを示す。

マルコフ政策クラス Π 上の最大化問題は

$$\begin{aligned} &\mathrm{Max} \quad P^{\pi}_{x_0}(r_0 \wedge \cdots \wedge r_{N-1} \wedge r_N \ge c) \\ &\text{s.t.} \quad (\mathrm{i})_n,\ (\mathrm{ii})_n \quad n = 0,1,\ldots,N-1 \end{aligned}$$

で表される。この閾値確率最大化問題に対しては、期待値問題に変換することなく、時刻 n で状態 $x_n(\in X)$ から始まる部分問題

$$\begin{aligned} &\mathrm{Max} \quad P^{\pi}_{x_n}(r_n \wedge \cdots \wedge r_N \ge c) \\ &\text{s.t.} \quad (\mathrm{i})_m,\ (\mathrm{ii})_m \quad m = n,\ldots,N-1 \end{aligned}$$

のマルコフ政策 $\pi = \{\pi_n, \pi_{n+1}, \ldots, \pi_{N-1}\} \in \Pi(n)$ にわたる最大値を $f_n(x_n)$ とする。ただし

$$f_N(x_N) \triangleq \phi(r_N(x_N)).$$

ここに ϕ は区間 $[c,\infty)$ の定義関数である。

このとき、次の関係式を得る。

補題 5.1 任意のマルコフ政策 $\pi = \{\pi_n, \ldots, \pi_{N-1}\}$ と任意の $x_n \in X$ に対して、

$$P^{\pi}_{x_n}(r_n \wedge \cdots \wedge r_N \ge c) = \begin{cases} \displaystyle\sum_{x_{n+1}\in X} P^{\pi'}_{x_{n+1}}(r_{n+1} \wedge \cdots \wedge r_N \ge c)p(x_{n+1}|x_n,u_n) & \textit{if} \quad r_n \ge c \\ 0 & \textit{otherwise} \end{cases} \tag{4}$$

が成り立つ。ここに

$$r_n = r(x_n,u_n), \qquad u_n = \pi_n(x_n), \qquad \pi' = \{\pi_{n+1}, \ldots, \pi_{N-1}\}.$$

補題 5.1 を多重和で表すと、次になる。

補題 5.2 任意のマルコフ政策 $\pi = \{\pi_n, \ldots, \pi_{N-1}\}$ と任意の $x_n \in X$ に対して、

$$\sum_{(x_{n+1},x_{n+2},\ldots,x_N)\in(*)} \sum\cdots\sum p_n p_{n+1} \cdots p_{N-1} = \begin{cases} \displaystyle\sum_{x_{n+1}\in X} \left[\sum_{(x_{n+2},\ldots,x_N)\in(*)} \sum\cdots\sum p_{n+1} \cdots p_{N-1} \right] p(x_{n+1}|x_n,u_n) & \textit{if} \quad r_n \ge c \\ 0 & \textit{otherwise} \end{cases} \tag{5}$$

が成り立つ。ただし

$$p_m = p(x_{m+1}|x_m, u_m),\ u_m = \pi_m(x_m)\ ; r_n = r(x_n, u_n),\ u_n = \pi_n(x_n).$$

また、$(*)$ は $r_n(x_n, u_n) \wedge \cdots \wedge r_N(x_N) \geq c$ を満たす $(x_{n+1}, \ldots, x_N) \in X \times \cdots \times X$ 全体にわたる多重和であり、$(\star)$ は $r_{n+1}(x_{n+1}, u_{n+1}) \wedge \cdots \wedge r_N(x_N) \geq c$ を満たす $(x_{n+2}, \ldots, x_N)$ 全体にわたっている。

したがって、上述の補題から後向きの再帰式が成り立つ：

定理 5.1

$$f_n(x) = \begin{cases} \underset{u;r(x,u)\geq c}{\mathrm{Max}} \sum_y f_{n+1}(y)p(y|x,u) & \textit{if}\ \ \exists\, u\ ;\ r(x,u) \geq c \\ 0 & \textit{otherwise} \end{cases} \qquad (6)$$

$$x \in X,\ \ 0 \leq n \leq N-1$$

$$f_N(x) = \begin{cases} 1 & \textit{if}\ \ r(x) \geq c \\ 0 & \textit{otherwise} \end{cases} \qquad x \in X. \qquad (7)$$

さて、式 (6) の最大（値に到達する）点の全体を $\pi_n^*(x)$ としよう。すなわち、

$$\pi_n^*(x) = \begin{cases} \text{Max に到達する } u; r(x,u) \geq c \text{ の全体} & \textit{if}\ \ \exists\, u\ ;\ r(x,u) \geq c \\ \text{任意の } u \in U & \textit{otherwise} \end{cases} \qquad (8)$$

$$x \in X,\ \ 0 \leq n \leq N-1$$

このようにして得られたマルコフ政策 $\pi^* = \{\pi_0^*, \pi_1^*, \ldots, \pi_{N-1}^*\}$ は最適である。

6　3－2－2モデル

ここでは3状態2決定2段モデルにおいて、最小型評価値が $c = 0.7$ 以上になる閾値確率を最大化する問題を考える：

$$\begin{array}{ll} \mathrm{Max} & P_{x_0}^{\mu}(\, r_0(U_0) \wedge r_1(U_1) \wedge r_2(X_2) \geq 0.7\,) \\ \text{s.t.} & \text{(i)}\ \ X_{n+1} \sim p(\,\cdot\,|x_n, u_n) \\ & \text{(ii)}\ \ u_0 \in U,\ u_1 \in U \end{array} \quad n = 0, 1$$

数値例として、Bellman & Zadeh([1]) の問題を考え、3つの方法で共通の最適解が得られることを示す。

$r_2(s_1) = 0.3 \quad r_2(s_2) = 1.0 \quad r_2(s_3) = 0.8$

$r_1(a_1) = 1.0 \quad r_1(a_2) = 0.6$

$r_0(a_1) = 0.7 \quad r_0(a_2) = 1.0$

$u_t = a_1$

$x_t \backslash x_{t+1}$	s_1	s_2	s_3
s_1	0.8	0.1	0.1
s_2	0.0	0.1	0.9
s_3	0.8	0.1	0.1

$u_t = a_2$

$x_t \backslash x_{t+1}$	s_1	s_2	s_3
s_1	0.1	0.9	0.0
s_2	0.8	0.1	0.1
s_3	0.1	0.0	0.9

6.1 再帰式

まず、再帰式 (9) を解いて、拡大マルコフ政策クラス $\tilde{\Pi}$ の中で最適値関数列 $\{u^0(x_0;\lambda_0),\ u^1(x_1;\lambda_1),\ u^2(x_1;\lambda_2)\}$ および最適政策 $\gamma^* = \{\gamma_0^*(x_0;\lambda_0),\ \gamma_1^*(x_1;\lambda_1))\}$ が求められる。これをまとめると、表1になる。

$$
\begin{aligned}
u^2(x_2;\lambda_2) &= \psi(\lambda_2 \wedge r_2(x_2)) \\
u^1(x_1;\lambda_1) &= \mathop{\mathrm{Max}}_{u_1} \sum_{x_2} u^2(x_2;\lambda_1 \wedge r_1(u_1))p(x_2|x_1,u_1) \\
u^0(x_0;\lambda_0) &= \mathop{\mathrm{Max}}_{u_0} \sum_{x_1} u^1(x_1;\lambda_0 \wedge r_0(u_0))p(x_1|x_0,u_0)
\end{aligned}
\tag{9}
$$

	$u^2(x_2;\lambda_2)$			$u^1(x_1;\lambda_1)$	$\gamma_1^*(x_1;\lambda_1)$	$u^0(x_0;0)$	$\gamma_0^*(x_0;0)$
$x_n \backslash \lambda_n$	0.6	0.7	1.0	0.7	1.0	0	
s_1	0	0	0	0.2 a_1	0.2 a_1	0.92	a_2
s_2	0	1	1	1.0 a_1	1.0 a_1	0.28	a_1, a_2
s_3	0	1	1	0.2 a_1	0.2 a_1	0.28	a_1

表1 拡大マルコフ政策クラスの最適解

次に、再帰式 (10) を解くと、原始政策クラス Π_p の中で最適値関数列 $\{w^0(x_0),\ w^1(x_0,u_0,x_1),\ w^2(x_0,u_0,x_1,u_1,x_2)\}$ および最適政策 $\hat{\mu} = \{\ \hat{\mu}_0(x_0),\ \ \hat{\mu}_1(x_0,u_0,x_1))\}$ が得られる。これは次の表2,3,4になる。

$$
\begin{aligned}
&w_2(x_0,u_0,x_1,u_1,x_2) = \psi(r_0(u_0) \wedge r_1(u_1) \wedge r_2(x_2)) \\
&w_1(x_0,u_0,x_1) = \mathop{\mathrm{Max}}_{u_1} \sum_{x_2} w_2(x_0,u_0,x_1,u_1,x_2)p(x_2|x_1,u_1) \\
&w_0(x_0) = \mathop{\mathrm{Max}}_{u_0} \sum_{x_0} w_1(x_0,u_0,x_1)p(x_1|x_0,u_0)
\end{aligned}
\tag{10}
$$

u_0	a_1		a_2	
$x_2 \backslash u_1$	a_1	a_2	a_1	a_2
s_1	0	0	0	0
s_2	1	0	1	0
s_3	1	0	1	0

表2 $\{w_2(x_0,u_0,x_1,u_1,x_2)\}$

x_0	$w_0(x_0)$	$\hat{\mu}_0(x_0)$
s_1	0.92	a_2
s_2	0.28	a_1, a_2
s_3	0.28	a_1

表4 $\{w_0(x_0)\ \ \hat{\mu}_0(x_0)\}$

x_1	$w_1(x_0,a_1,x_1)$	$\hat{\mu}_1(x_0,a_1,x_1)$	$w_1(x_0,a_2,x_1)$	$\hat{\mu}_1(x_0,a_2,x_1)$
s_1	0.2	a_1	0.2	a_1
s_2	1	a_1	1	a_1
s_3	0.2	a_1	0.2	a_1

表3 $\{w_1(x_0,u_0,x_1)\ \ \hat{\mu}_1(x_0,u_0,x_1)\}$

このとき、γ^* から生成される一般政策 σ^* は $\hat{\mu}$ から生成される一般政策 $\hat{\sigma}$ に一致し、これはまた次の多段確率決定樹表によっても得られることがわかる。

さらに、再帰式(11)を解くと、マルコフ政策クラス Π の中で最適値関数列 $\{f_0(x_0), f_1(x_1), f_2(x_2)\}$ および最適政策 $\pi^* = \{\pi_0^*(x_0), \pi_1^*(x_1)\}$ が得られる。これをまとめると、表5になる。

$$
\begin{aligned}
f_2(x_2) &= \begin{cases} 1 & \text{if } r_2(x_2) \geq 0.7 \\ 0 & \text{otherwise} \end{cases} \\
f_1(x_1) &= \underset{u_1; r_1(u_1) \geq 0.7}{\text{Max}} \sum_{x_2} f_2(x_2) p(x_2 | x_1, u_1) \\
f_0(x_0) &= \underset{u_0; r_0(u_0) \geq 0.7}{\text{Max}} \sum_{x_1} f_1(x_1) p(x_1 | x_0, u_0)
\end{aligned}
\tag{11}
$$

x_n	$f_2(x_2)$	$f_1(x_1)$	$\pi_1^*(x_1)$	$f_0(x_0)$	$\pi_0^*(x_0)$
s_1	0	0.2	a_1	0.92	a_2
s_2	1	1.0	a_1	0.28	a_1, a_2
s_3	1	0.2	a_1	0.28	a_1

表5 マルコフ政策クラスの最適解

マルコフクラスでの最適政策 π^* は、拡大マルコフクラスでの最適政策 γ^* から生成された一般最適政策 σ^* および原始政策クラスでの最適政策 $\hat{\mu}$ から圧縮された一般最適政策 $\hat{\sigma}$ に(マルコフ政策として)一致している：

$$\sigma^* = \hat{\sigma} = \pi^*.$$

6.2 多段確率決定樹表

多段確率決定樹表は、問題のデータを過程の進行状況に応じて配列し、あらゆる可能な経路とその評価値・確率を図示し、各段における最適決定の選択を明示している。この意味では列挙法の解構成を与えている。しかし、最適解に至るまでは動的計画法の再帰式を解く順に構成されている。この樹表ではあらゆる型の評価関数の期待値最適化が解かれる。

次頁の樹表（図1）では、3-2-2型（3状態2決定2段）最小型モデルに対して次のように簡略化している(数値自体は Bellman and Zadeh (1970) のデータ)：

履歴 $= x_0 \;\; r_0(u_0)/u_0 \;\; p_0 \;\; x_1 \;\; r_1(u_1)/u_1 \;\; p_1 \;\; x_2 \;\; r_2(x_2)$

ただし $p_0 = p(x_1 \mid x_0, u_0), \quad p_1 = p(x_2 \mid x_1, u_1)$

最小型評価 $=$ 最小型評価値 $= r_0(u_0) \wedge r_1(u_1) \wedge r_2(x_2)$

経路確率 $= p_0 p_1$

閾値確率 $= \psi(r_0(u_0) \wedge r_1(u_1) \wedge r_2(x_2)) p_0 p_1$

ただし $\psi(y) = 1_{[2.5,\infty)}(y)$

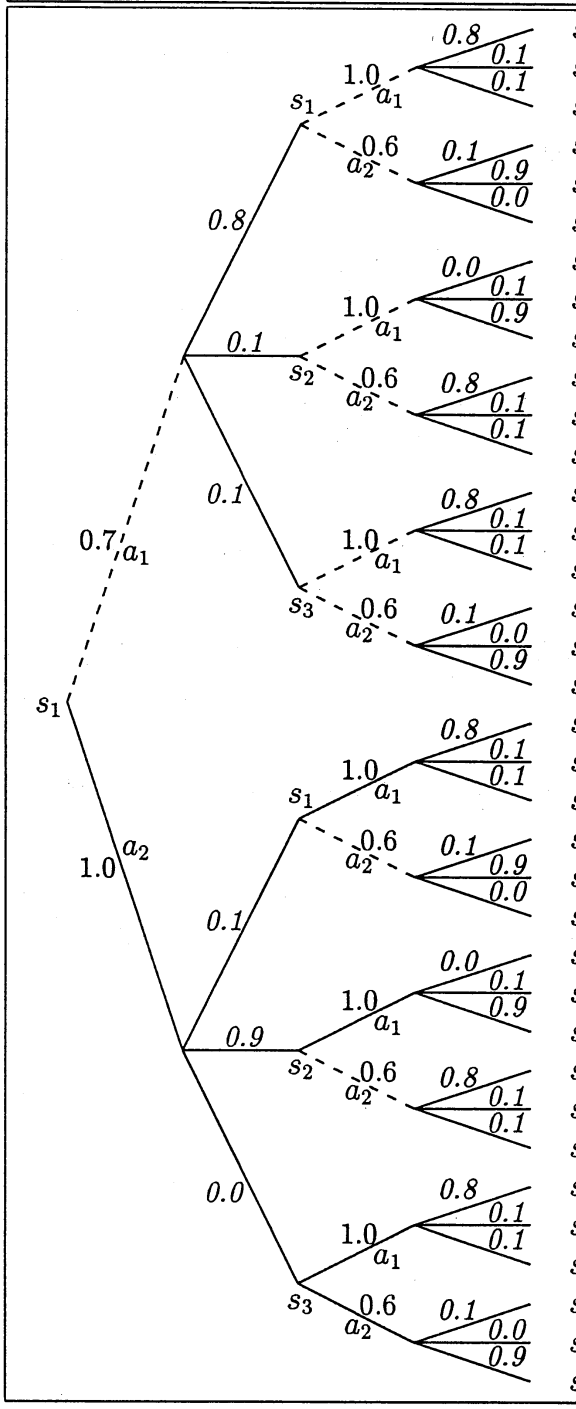
部分確率 $= \sum_{x_2} \psi(r_0(u_0) \wedge r_1(u_1) \wedge r_2(x_2)) p_0 p_1$

全確率 $=$ 全体確率 $= \sum_{x_1} \sum_{x_2} \psi(r_0(u_0) \wedge r_1(u_1) \wedge r_2(x_2)) p_0 p_1.$

イタリック体は確率を、ボールド体は上下の確率のうち大きい方を選択したことを表す。特に、履歴の欄では5つの数値 $r_0 = r_0(u_0)$, p_0, $r_1 = r_1(u_1)$, p_1, $r_2 = r_2(x_2)$ のみを記している。

$$V^0(s_1) = \max_{\mu} P^{\mu}_{s_1}(\, r_0(U_0) \wedge r_1(U_1) \wedge r_2(X_2) \geq 0.7\,)$$

図 1：状態 s_1 からの 2 段確率決定樹表



履歴 x_0	r_0/u_0	p_0	x_1	r_1/u_1	p_1	x_2	r_2	最小型評価	経路確率	閾値確率	部分確率	全確率
s_1	a_1	0.7	s_1	a_1	0.8	s_1	0.3	0.3	*0.64*	0		
						s_2	1.0	0.7	*0.08*	0.08	**0.16**	
						s_3	0.8	0.7	*0.08*	0.08		
				a_2	0.1	s_1	0.3	0.3	*0.08*	0		
						s_2	1.0	0.6	*0.72*	0	0	
						s_3	0.8	0.6	*0.0*	0		
			s_2	a_1	0.1	s_1	0.3	0.3	*0.0*	0		
						s_2	1.0	0.7	*0.01*	0.01	**0.1**	
						s_3	0.8	0.7	*0.09*	0.09		0.28
				a_2	0.9	s_1	0.3	0.3	*0.08*	0		
						s_2	1.0	0.6	*0.01*	0	0	
						s_3	0.8	0.6	*0.01*	0		
			s_3	a_1	0.1	s_1	0.3	0.3	*0.08*	0		
						s_2	1.0	0.7	*0.01*	0.01	**0.02**	
						s_3	0.8	0.7	*0.01*	0.01		
				a_2	0.9	s_1	0.3	0.3	*0.01*	0		
						s_2	1.0	0.6	*0.0*	0	0	
						s_3	0.8	0.6	*0.09*	0		
s_1	a_2	1.0	s_1	a_1	0.8	s_1	0.3	0.3	*0.08*	0		
						s_2	1.0	1.0	*0.01*	0.01	**0.02**	
						s_3	0.8	0.8	*0.01*	0.01		
				a_2	0.1	s_1	0.3	0.3	*0.01*	0		
						s_2	1.0	0.6	*0.09*	0	0	
						s_3	0.8	0.6	*0.0*	0		
			s_2	a_1	0.9	s_1	0.3	0.3	*0.0*	0		
						s_2	1.0	1.0	*0.09*	0.09	**0.9**	
						s_3	0.8	0.8	*0.81*	0.81		**0.92**
				a_2	0.1	s_1	0.3	0.3	*0.72*	0		
						s_2	1.0	0.6	*0.09*	0	0	
						s_3	0.8	0.6	*0.09*	0		
			s_3	a_1	0.0	s_1	0.3	0.3	*0.0*	0		
						s_2	1.0	1.0	*0.0*	0.0	**0**	
						s_3	0.8	0.8	*0.0*	0.0		
				a_2	0.9	s_1	0.3	0.3	*0.0*	0		
						s_2	1.0	0.6	*0.0*	0.0	**0**	
						s_3	0.8	0.6	*0.0*	0		

References

[1] R.E. Bellman and L.A. Zadeh, *Decision-making in a fuzzy environment*, Management Science, **17**(1970), pp.B141-B164.

[2] M. Bouakiz and Y. Kebir, *Target-level criterion in Markov decision processes*, Journal of Optimization Theory and Applications, **86**(1995), pp.1-15.

[3] 岩本 誠一, 多段確率決定樹表について, 日本OR学会秋季研究発表会アブストラクト集, 1999, pp.58-59.

[4] 岩本 誠一, 確率最適化における再帰式と確率決定樹表, 研究集会「不確実、不確定性の下での数理的決定理論」, 京大数理研講究録 1132, 2000, pp.15-23.

[5] S. Iwamoto, *Maximizing threshold probability through invariant imbedding*, Ed. H.F. Wang and U.P. Wang, Proceedings of THE EIGHTH BELLMAN CONTINUUM, Hsinchu, ROC, Dec.2000, pp.17-22.

[6] S. Iwamoto, *Fuzzy decision-making through three dynamic programming approaches*, Ed. H.F. Wang and U.P. Wang, Proceedings of THE EIGHTH BELLMAN CONTINUUM, Hsinchu, ROC, Dec.2000, pp.23-27.

[7] S. Iwamoto and T. Fujita, *Stochastic decision-making in a fuzzy environment*, J. Operations Res. Soc. Japan, **38**(1995), pp.467-482.

[8] S. Iwamoto, T. Ueno and T. Fujita, *Controlled Markov chains with utility functions*, Proc. of The International Workshop on Markov Processes and Controlled Markov Chains, Changsha, China, 1999, to appear.

[9] 植野 貴之, 岩本 誠一, 最小型評価系の閾値確率制御, 日本OR学会秋季研究発表会アブストラクト集, 2000, pp.124-125.

[10] C. Wu and Y. Lin, *Minimizing risk models in Markov decision processed with policies depending on target values*, Journal of Mathematical Analysis and Applications, **231**(1999), 47-67.